

---

# CSRの不足した企業の特徴分析

2016年度グループ演習4班 小野口孟 藤村美月 金雪飛 徐千里

アドバイザー教員：掛谷英紀

---

## 1. 背景・目的

近年、企業の社会的責任(CSR: Corporate Social Responsibility)が注目されている。CSRは、納税や法令遵守はもちろんのことながら、公正な事業活動の推進、地域課題への取り組みなど様々なことを指す。しかし国内では三菱自動車の燃費試験不正問題や電通の広告費不正請求など、企業の不祥事が後を絶たない。このような企業の不祥事は国内だけでの問題ではない。世界的な問題の一つに企業のタックスヘイブンを利用した納税逃れがある。この行為は以前から問題視されていたが、企業のタックスヘイブンの利用実態がわからず野放しにされていた。しかし2016年に国際ジャーナリスト連合(ICIJ: International Consortium of Investigative Journalists)によってタックスヘイブンを利用していた企業リスト(パナマ文書)が発表された。パナマ文書には企業名はもちろんのことながら、世界的著名人の名前も多く掲載されており、世界に衝撃を与えた。そして、そのパナマ文書には日本企業も掲載されており、日本企業のタックスヘイブンの利用実態も明らかとなった。

タックスヘイブン利用の最大の問題は、国の税収が減少することである。その影響は社会保障費の減少や増税といった形で国民に及ぶ<sup>(1)</sup>。そのためタックスヘイブンの利用は法律的にはグレーゾーンであるが、CSRという観点では明らかに問題がある。そこで、本研究はそうした行為に及ぶ企業の風土や文化を明らかにすることを目的とする。

企業風土や企業文化の分析手法の一つとして、自然言語処理による経営トップのメッセージ分析がある。佐藤ら<sup>(2)</sup>は自然言語処理と機械学習を用いて社長のメッセージをもとにその企業の平均勤続年数が長いかどうかを推定する手法を提案して

いる。クロスバリデーションの正解率は高くないものの、平均勤続年数が長い企業と短い企業に特徴的な表現は見出されている。本研究でも、この分析手法を用いる。

パナマ文書に掲載された日本企業のリストを見ると、同じ業界内でも掲載されている企業とそうでない企業が存在していることがわかる。そこで本研究では、パナマ文書と以前に発表されていたオフショアリークスに掲載された日本企業と掲載されていない同業他社の社長のインタビュー記事をもとに、掲載企業と非掲載企業の特徴分析を行う。また、四季報などに掲載されている企業データ、および過去の不祥事件数などからの分析も行う。そしてタックスヘイブン利用企業とそうでない企業の風土や文化の違いを明らかにすることを目的とする。

## 2. 分析手法

### 2.1 自然言語処理

インターネット上で収集したオフショアリークスとパナマ文書の掲載企業と非掲載企業の経営トップのインタビュー記事に特徴の違いがあるかどうかを、自然言語処理を用いて分析を行う。以下に分析手法を示す。まず、インタビュー記事のテキストから形態素解析ツール(ChaSen)を使って、それぞれ単語ごとに分割し、その中から特定の品詞を選んで素性データとする。次にその素性データを、最大エントロピー(ME: Maximum Entropy)法を用いた機械学習にかけることでカテゴリ分類を行う。最大エントロピー法は、スパースなデータを得意とする統計的機械学習アルゴリズムで、文書分類において頻繁に使われている手法の一つである。ME法の学習にはmaxentソフトウェアを

使用した。

機械学習時のクロスバリデーションでは、業種間で社長のインタビュー記事の特徴に違いがある可能性を考慮し、同業種の企業を対にしてクロスバリデーションを行った。クロスバリデーションとは、対象とするデータを任意の数のグループに分割し、その中の1グループをテストデータ、他のグループを機械学習の学習データとするものである。例えば、10の記事データに対して、分割数5でクロスバリデーションを行うと、2記事がテストデータに、8記事が学習データとなる。そして、8記事をもとに2記事をカテゴリ毎に分類する。これを全組み合わせ、つまり5回行うことで全ての記事のカテゴリ分類の正答率を評価する。

機械学習に使用する経営トップのインタビュー記事はオフショアリークスとパナマ文書に掲載された日本企業と掲載されていない同業他社に分けて社長のインタビュー記事を48件ずつ合計96件収集した。

## 2.2 統計分析

オフショアリークスとパナマ文書に掲載された企業と掲載されていない同業他社の間に、企業統計データの特徴的差異があるかを調べるため、統計分析ソフト SPSS を用いた T 検定を行う。今回分析で用いる企業には複数の業界に属するものが混在しており、業界の相違による影響を避けるため「対の T 検定」を行う。対の T 検定を用いて対応する同じ業界の2社ごとの差を集計し、それらの差の平均と分散から有意差があるかどうかを判断する。

分析に用いたデータ項目は、東洋経済(2016)<sup>3)</sup>より修士・大卒採用数(名)、3年後離職率(%), 有給消化年平均(日)、平均年収(万円)、従業員(人)、月平均残業時間(時間)、平均勤続年齢(歳)、平均勤続年数(年)の8項目を抽出した。また各企業のHPより過去5年間の売上高、社長平均在任期間を抽出した。そのほか企業の広告宣伝費・売上高広

告宣伝費率<sup>4)</sup>、不祥事事件についても分析する。

## 3. 分析結果

### 3.1 インタビュー記事の分析結果

インタビュー記事から形態素解析で、名詞および、形容詞と動詞の基本型を素性データとして抽出した。そしてそのデータから機械学習をし、48分割のクロスバリデーションをした結果、正解率は57.3%となった。今回のデータは、掲載と非掲載の2カテゴリであることからランダムに答えても50%の正解率は保証される。そのためこの正答率は十分高いとは言えないが、データ数が96と少ないことを考えると、一定の特徴を掴むことはできていると考えられる。

次に、各カテゴリの頻出上位素性35とその素性のインタビュー記事内での出現頻度を図1に示す。図1からオフショアリークス、パナマ文書掲載企業と非掲載企業の違いが見られる。掲載企業の社長が「グループ」、「運営」、「子会社」、「会議」など会社やグループ企業に関する内向きの言葉を多く使用しているのに対し、非掲載企業の社長は、「お客様」、「価格」、「商品」などの顧客を中心とした外向きの言葉を多く使用している。さらに、後者で特異に多く見られる素性として、「向け」、「市場」という単語がある。記事内での使われ方を見ると、「向け」は「消費者向け」、「高齢者向け」などの表現が見られる。「市場」も、「新興市場への進出」、「市場のニーズ」などのように商品などの市場動向を意識している発言が多い。これらのことから、非掲載企業は掲載企業と比較してより顧客を重視している特徴があると考えられる。

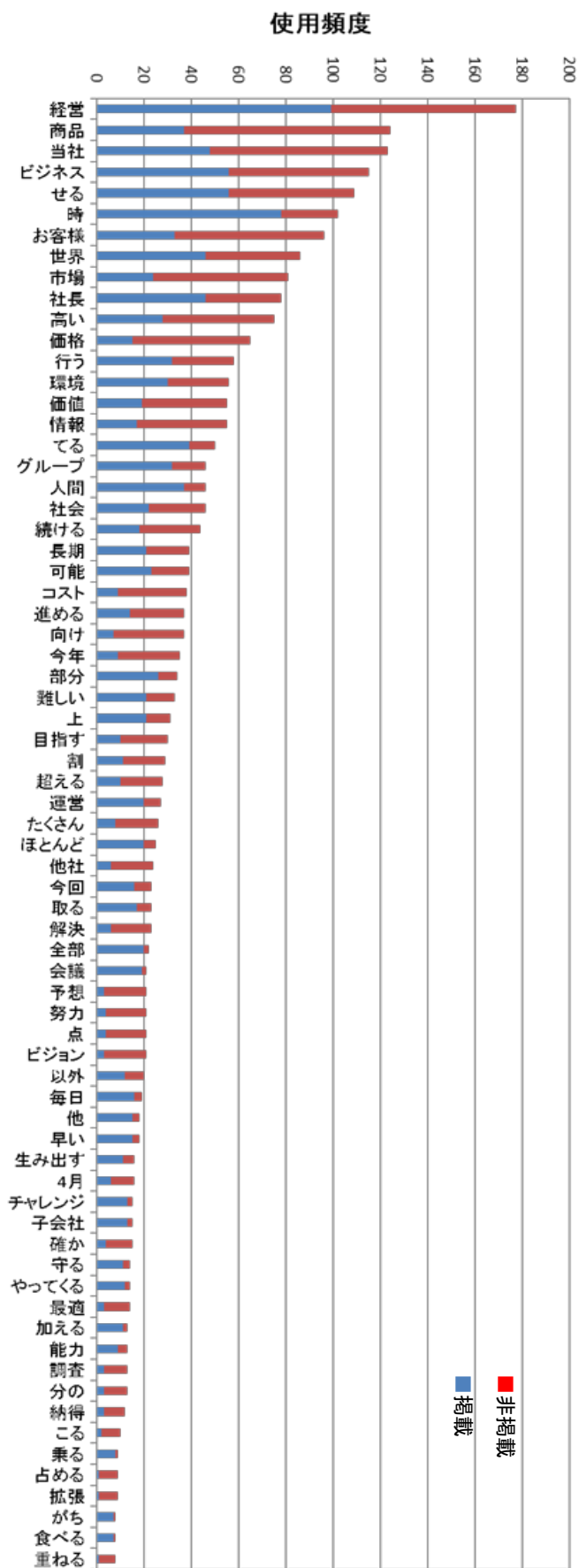


図1 各カテゴリ上位素性の使用頻度

### 3.2 企業データの分析結果

#### (1) 四季報・売上高による企業分析

オフショアリークスとパナマ文書に掲載された企業と非掲載企業との違いを明らかにするため、四季報から抽出した8項目と過去5年間の売上高をT検定により分析した。結果を表1に示す。

分析対象である企業が少ないため、分析対象としたすべての項目において5%水準の有意な差は見られていない。ただし、3年後離職率に関してはパナマ文書非掲載企業の方がパナマ文書掲載企業より離職率が低い傾向が見られている(p = .107)。このことから、パナマ掲載企業は社員の入れ替わりが多く、企業風土として何らかの問題を抱えている可能性はある。

表1 掲載企業と非掲載企業の対のT検定結果

	パナマ文書 掲載・非掲載	度数	平均値	標準偏差	t値	有意確率 p値
修士・大卒 採用数(名)	掲載企業	23	133	147	1.31	.205
	非掲載企業	23	103	142		
3年後 離職率(%)	掲載企業	13	7.5	7.1	1.74	.107
	非掲載企業	13	4.1	4.0		
有給消化 日数(日/年)	掲載企業	22	10.9	3.7	1.32	.200
	非掲載企業	22	9.9	2.9		
平均年収 (万円)	掲載企業	22	851.8	376.3	0.28	.782
	非掲載企業	22	833.0	205.0		
従業員(人)	掲載企業	25	4583	4038	-0.57	.573
	非掲載企業	25	5438	7324		
月平均残業時間 (時間)	掲載企業	19	23.3	8.4	-0.41	.688
	非掲載企業	19	24.0	8.6		
平均年齢 (歳)	掲載企業	24	40.4	1.9	1.25	.224
	非掲載企業	24	39.9	2.2		
平均勤続年数 (年)	掲載企業	24	15.3	3.1	0.46	.649
	非掲載企業	24	15.0	2.5		
過去5年間の平均 売上高(百億円)	掲載企業	33	228.5	352.6	1.35	.185
	非掲載企業	33	146.2	241.7		

#### (2) 広告宣伝費による企業分析

東洋経済オンラインの広告宣伝費の上位50社ランキング<sup>(4)</sup>中に、オフショアリークス・パナマ文書掲載企業とその競合他社が何社ランクインしているか、また売上高と広告宣伝費の比の違いを評価した。

掲載企業でランクインしたのは6社、非掲載企業では10社ランクインしている。掲載企業とその競合他社の両方ランクインしているのは4組であ

った。そして掲載企業のみがランクインしたのは 2 社、非掲載企業のみがランクインしたのは 6 社であった。このことから非掲載企業の方が広告にお金を掛けている企業が多いと考えられる。これ自然言語処理での結果である、非掲載企業の方が企業の外側を重視しているという点と一致し、経営者が会社の商品や会社自体を外部に発信しようとする意識が強いと考えられる。

次に、上位 500 社にランクインした掲載企業と非掲載企業の売上高と広告宣伝費の比の平均はそれぞれ 2.85 %と 2.98 %であり、両者に差はみられなかった。しかし、今回は広告宣伝費の上位 500 社ランキングにランクインしている企業のみを対象としていることから、企業データを増やすことで差が出てくる可能性がある。

### (3) 社長の在任期間による企業分析

同じ人物が長く社長を続けている企業は、隠し資産を形成しやすい可能性があるとの仮定の下、パナマ文書掲載企業と非掲載企業の社長の平均在任期間を比較した。まず、各企業の創業者から歴代の社長の在任期間を調査し、調査できたパナマ文書掲載企業 26 社、非掲載企業 19 社の平均値を求めた。その結果、社長在任期間の平均値はパナマ文書掲載企業 M=11.4(年)、非掲載企業 M=12.2(年)であり、大きな差は見られなかった。

### (4) 企業の不祥事事件による企業分析

パナマ文書掲載企業は非掲載企業より脱税などの不祥事事件の数が多いと仮定し、過去 10 年間の企業の不祥事事件を調査した。ニュースサイトから、脱税や金銭関係の不祥事により新聞記事になった企業を抽出、パナマ文書掲載企業か非掲載企業に分けて整理をした。結果、パナマ文書掲載企業は 5 件、非掲載企業は 2 件であった。不祥事の内容は、パナマ文書掲載企業では税金の申告漏れや所得隠し、架空請求とあり、非掲載企業は、賃金未払いや税金の申告漏れがあった。

## 4. まとめ

本研究では、オフショアリークス、パナマ文書掲載企業と非掲載企業の企業風土や文化を明らかにすることを目的とし、自然言語処理による社長インタビュー記事の分析と企業データ分析の調査を行った。

自然言語処理では、正解率は 57.3 %と高い数値は得られなかったが、上位素性からは各カテゴリの特徴を得ることができた。その結果、掲載企業は企業の内側を重視し、非掲載企業は企業の外側を重視しているといえる。

企業データの T 検定では、四季報の企業データ、売上高からパナマ文書掲載企業と非掲載企業の違いはないことが示された。ただし、3 年後離職率に関しては非掲載企業の方がパナマ文書掲載企業より低い可能性が示唆された。また、社長の平均在任期間を比較したところ、掲載企業と非掲載企業間には大きな差は見られなかった。つまり、社長の在任期間の長短では掲載企業か否かを判断することは難しいといえる。ただし、社長よりもオーナーや会長が影響力を持つ企業もあり、株主比率や会長の在任期間なども分析すると、何らかの差が見られる可能性がある。

本研究では、オフショアリークス、パナマ文書掲載企業と非掲載企業の企業風土や文化に一定の特徴を見出すことに成功したが、分析対象とした企業数が少ないため、まだ十分信頼性の高い結果は得られていない。今後は分析対象とするデータ数を増やし、分析精度向上を行っていく必要がある。

## 5. 参考文献

- (1) 渡辺哲也(2016)『パナマ文書 タックスヘイブン狩りの衝撃が世界と日本を襲う』第二刷株式会社徳間書店.
- (2) 佐藤裕也、掛谷英紀(2015)「企業のトップメッセージに基づく企業風土の特徴分析」言語処理学会第 21 回年次大会発表論文集、pp.473-476
- (3) 東洋経済新報社(2016)『就職四季報 2017 年版』大日本印刷.
- (4) 東洋経済(2016)「『広告宣伝費』トップ 500 社ランキング」[http://toyokeizai.net/articles/-/132053\(2016/10/16\)](http://toyokeizai.net/articles/-/132053(2016/10/16))